

Prediction of Job Applicant Salary and Designation using Machine Learning

Ashok Kumar S, Santhosh V, Saranyadevi S, Vikranth Krishna E



Abstract: This system is aimed toward developing a central recruitment process system for the HR Group for an organization. This system is fully supported by recruiting a candidate for a job supported their performance within the interview and their test scores are analyzed to predict their basic salary once a year. Some features of this system are getting to be creating vacancies for candidates, storing application data, and Scheduling interviews, Interview process initiation, storing Interview results for the candidate and eventually Hiring of the candidate. Reports could even be required to be generated for the use of the HR group. This project "Hiring Dataset Analysis" could also be a windows application during which the candidate details store by admin then attends the drive. Supported the results of the qualification the job seekers are getting to be shortlisted. For freshers, the drive is getting to be conducted at some venue after the shortlisting of the chosen candidate. The tiny print of the particular company drive, venue & Date of the examination are getting to be made available to them through the online site. Functionality and Features of the candidate profile, posting new recruitment programs by company, publishing new carrier guidance placement programs by Admin, Viewing Resumes by company, Viewing payment by candidate, Company profile management.

Keywords: Salary prediction, e-Recruitment, Job Market, Machine Learning, K-means clustering.

I. INTRODUCTION

Consultancy firms are considered desirable places to work, with students curious about the promise of interesting, challenging and diverse work also because of the chance for early and substantial career progression. But do you know what consulting actually is? What types of consultancy employers are out there? And what exactly do these roles and corporations require from you? Fear not, we are here to answer all of those questions and more some market enhancements, changes in management,

optimizing growth or recognizing a competitive advantage are usually taken to consultants. Sometimes the field of work of a consultant is extremely varied and consultants are used in virtually every sector. At the beginning of your career, some main consultancy areas are business planning, manufacturing and business services, financial and management controls, HR, marketing, environmental management, quality management, IT. Consultants in all industry fields have the following qualities in common: critical thinking, good communication skills, the ability to work in a team and effective problem-solving capabilities.

II. LITERATURE SURVEY

In this paper^[9] they demonstrated that there are vast data in the medical industry out there to obtain information from large amounts of data using an analytical method. A true data collection has been taken from SGPGI during this article. Real-time data sets are often interconnected with problems such as incomplete values, higher dimension values, and interference, etc. that are not effective for the entire classification. Hence clustering is that the complementary data analytics approach. The main objective of this paper is to develop a unique technique focused on K-mean clustering. The experimental results show that k-means clustering the algorithm features an outstanding result on dataset that is actual relative to simple k-means clustering algorithms and offers an improved outcome to the actual-world problem.

Within this paper^[11] they suggested a method called the Student Performance Analysis Program to keep track of the leads of the student to a specific university. The proposed development provides a framework that forecasts the students 'success in support of their evaluating and design informed by tests. The proposed solution provides the prediction of student success through the principles developed by a processing technique. The data mining strategy used in this project is classification, which defines the grades of the students sponsored by the scholars

III. EXISTING SYSTEM

The existing system may be a manual one using which the Hiring Dataset Analysis agent cannot maintain the effect by sharing across a different job with proper security and can't track details easily. It does not provide proper coordination between the different candidates of the company. It does not allow the candidate to check the placement status of his file properly, which leads to customer dissatisfaction.

Revised Manuscript Received on April 30, 2020.

* Correspondence Author

Ashok kumar S*, Department of Information Technology, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India. Email: ashokkumar@siet.ac.in

Santhosh V, Department of Information Technology, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India. Email: santhoshram283@gmail.com

Saranyadevi S, Department of Information Technology, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India. Email: saranyadevi1999@gmail.com

Vikranth Krishna E, Department of Information Technology, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India. Email: vikranthkrishna99@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

The application is not updated with the candidate job application count based on consultancy payment. The error occurring in Applications is not solved. The auto calculation takes more time to load. The save and update operation is not combined. The report generation is very slow and also unaligned. The admin dashboard is difficult to understand.

IV. PROPOSED SYSTEM

In the Proposed System the application is fully developed using Python and MySQL. The desktop automated system with computer-based architecture can support issues

The application is updated with the candidate job application count based on consultancy payment. Application error occurring is fully solved.

The auto calculation takes less time to load. The save and update operation is combined with binding navigator. The report generation is very fast and also aligned. The admin dashboard is easy to understand.

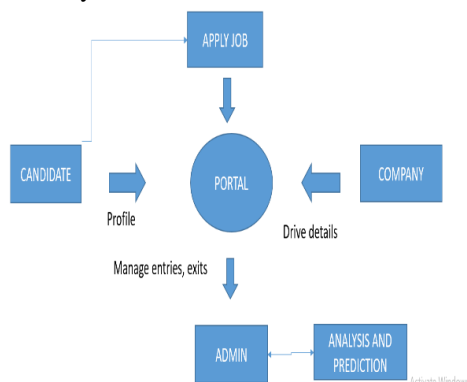


Fig 4.1 System Architecture

V. MODULE DESCRIPTION

A. Admin Module

This module is accessed by admin only. Admin can log in using the admin id and password. This admin panel is employed to vary (or) update any data during this. Because, for a security purpose data maintenance authority given to admin, he can only make any changes during this using login id and password.

B. Company Module

Within this module section of the Consultants application, the business information page will be displayed when we click on the "business" MDI Form. Now companies may register their data. Upon registration the data will be stored in a MySQL file and a Unique Id I.e. company Id will be issued to them. The Company -id can be used by clicking the "Delete" button on the home page for further use as for the deletion of the company.

C. Open Jobs Module

When the user clicks the "Open Task" hyperlink in this section, the company's tiny print of the work requirement page gets shown. Just a few corporate employees will be able to access this work requirement page only by entering a specific username and password. If an incorrect username or password

is given, the job requirement page cannot be accessed and thus the server will ask at most 3 times for correct username and password. The server will then shut down the order. On this tab, corporate staff may enter the specific specifications for their organization when entering the correct username and password. These specifications will be stored in the Database. A single work identifier is allocated for each specific job requirement, i.e. Job-id is automatically generated. Therefore, the task-id can be used to uninstall the work after the task criteria have been met by clicking on the "Uninstall" button on the alternative tab. This choice to delete gets clarified later. Here we use XML as a backend, rather than using MySQL which increases the speed of the application and reduces costs compared to MySQL.

D. Apply Jobs Module

In this section, when the user enters the "Apply Job" link, the details of the applicant's career field and experience site will come up. Upon entering the information, and submitting the details. Quite a certain organizational workers will be able to access this job requirement page by entering common login details only. If an incorrect username or password is received, the candidate request page can not be accessed, and the server will request the appropriate user name and password for a max of three occasions.

Then, the server will prohibit the order. After entering the appropriate login details, the list of candidates who meet the requirements can be displayed.

In the module portion, when the user clicks "Find a Job," the details of the candidate's skilled region, the list of needed companies will be displayed by the database search. It will also work as an employment investigation engine. The applicants will then apply for the position they have available.

In the unit section, the candidate can delete their portfolio from the database by simply placing the applicant-id (produced while submitting a portfolio in choice 1) if the applicant does not need the work. Remove choice can be found on the main page.

The business individuals may delete the work requirement of the business from the database as part of the unit by inputting the job-id (produced while entering the work in the business depends on the payment of the individual) once the organization reaches the job requirement. The Delete option can be found on the main page.

VI. METHODOLOGY

To obtain useful information into the online job recruitment for IT professionals, we try comparing different approaches and machine learning models. The approach reflects the best literature and business standards covering different phases:

- Data collection: A web sitemap based in Python is developed to analyze and collect the essential website information.
- Data cleaning: posts with lost values are eliminated, and potential data layout disputes (e.g. text encoding) are fixed.

- Manual feature engineering: inconsequential characteristics are thrown away, and other features are formalized by exploiting domain knowledge (e.g. transformed into mathematical features).
- Description of the data set: statistical tools and simple models are used to give a preliminary and compact description of the data.
- Automatic feature selection: functional selection algorithms are used to choose the most helpful features about the output class in a dataset.
- Model selection: a search to find the optimal hyper-parameters for a range of well-known machine learning models is performed.
- Model training and validation: train and cross-validate the selected models to check the classifiers that best characterize the data and can predict the performance variables.
- Comparison model: For standard scores and curves such as classification precision, each model is compared to the other.
- Using raw data sets we can predict the salary using k-means, linear Regression, Regularity linear regression. In this section, the k-means algorithm is used to predict the designation and salary for a candidate.

A. K-means Clustering

Clustering algorithms can group a set of samples according to their distances with numerical attributes. Instead of using all features within the dataset, we focus on sub-sets of relevant variables that divide job posts into groups based on their work parameters and ability requirements. K-means algorithm is:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|X_i^{(j)} - C_j\|^2$$

Where, $\|X_i^{(j)} - C_j\|^2$ can be a distance measure chosen between an information point and thus the cluster center, it is an indication of the n data points space from their respective cluster centers.

The first grouping scheme is based upon salary-. We have shown before that there are four with major linear dependency on the salary offered, namely Experience, Project. The use of these characteristics is related to their possible effect on salary. As a result, nine clusters divide the deals by their pay levels. Table 3 displays the value of each cluster's cancroids, sorted by the average salary. Experience is recorded as average years, while the number of posts in each category is expressed as Project Leader, Permanent Contract, and IT Architect. The resulting groups display linear increases in most characteristics, indicating, as predicted, that salaries and job conditions are mainly improving with experience. In addition, the number of permanent positions rises dramatically from early-career employment (low salaries, less than two years of average requirement experience) to existing professionals (more than 3 years of experience, higher wages), indicating that early-career movement is encouraged by businesses as experience tends to shape both wages and permanent contract probabilities.

Aiming to organize job posts into a skill-driven system, we replicate the k-means clustering experiments using only the skill keywords as settings. We get 5 clusters of skill-oriented posts, as well as the top ten most common skills within each group's posts. This separation scheme shows five well-defined market-requested profiles, where different skills offer access to different positions. In this scenario, there are 2099 posts in the first line, the fourth 679, the third 490, the fifth 432 and the second 258. Each of the groups has different patterns. The first cluster collects many competencies usually relevant to back-end developers and other non-web software, such as Oracle, SQL, PHP or SAP. The second cluster includes the required skills for a system administrator such as Linux, Windows, VMware, or Network.

The third and fourth clusters contain primarily skills needed by both. Net developers and Java developers. The fifth cluster corresponds to front-end developers of the most common technologies. From this viewpoint, it is clear that at Tecnoempleo, IT and technology are the skills most requested by recruiters. In addition, there are no major variations in terms of both income and experience, with both classes needing roughly two years of average experience. This demonstrates the presence of homogeneity across profiles, that is, no emerging and highly demanded skill sets appear to be present.

VII. RESULT

Depends on the wage We have shown that there are four features with substantial linear dependency on the salary offered. Let's lead, Perm. Contract Architect and IT. Apart from the strong, these four characteristics are those with the highest salary relation. We apply k-means clustering to the aforementioned features and the salary variable in order to find groups of job posts containing similar profiles. The use of these features is associated with their potential impact on salary. In terms of most features, the corresponding groups show linear rises, indicating, as predicted, that wages and working conditions are mainly improved with practice. In addition, the number of permanent positions rises dramatically from early-career workers (low pay, less than two years of average requirement experience) to existing professionals (further over 3 years experience, higher wages), indicating which early-career movement is encouraged by businesses, as knowledge tends to form both pay and lasting contract chances.

Prediction result

Applicant Name	Credit Score	Selection Result	Predicted Salary	Predicted Designation
Saranya	9	Selected	35,000	Software Developer
Arvind	8	Selected	30,000	Software Testing
Santhosh	7	Selected	25,000	Junior Software Developer
Swathi	7	Selected	22,000	Tecnical Support

Fig 7.1 Prediction Result

To enhance wage estimation accuracy, we propose the question as a classified function with four sections corresponding to small, scale-low, medium-moderate, and large pay scales. We are using the Python library implementation. Especially we hire multiple pipelines consisting of three key stages: Information standardization at intervals [0,1];, automated function choice (essential) that could be useful in decreasing dimension far beyond the product of our automatic non-processing function, grouping according to a particular model categorized by the better parameter specification as obtained by implementing a prior matrix search method.

VIII. CONCLUSION

The system has been developed with a lot of care that it's freed from errors and at the identical time, it's economical and fewer long. The necessary factor is that the system is strong. I even have tried our uttermost to create positioning as dynamic as attainable. The entire system is documented and may be simply understood by the end-users. It has been designed and runs to satisfy the wants and desires of the organization and the end-users. Also, provision is provided for future developments within the system. The complete system is secured. This online system is approved and enforced presently.

REFERENCES

1. Iftekhar Naim, M. Iftekhar Tanveer, Daniel Gildea, Mohammed Ehsan Hoque, "Automated job interview prediction and appraisal performance: the role of what you say and how you say it," Automatic Face and Management Identification (FG) 2015 11th Symposium and Workshops on IEEE, vol. 1, pp. 2, 2015.
2. Sowmya Rasipuram, Rahul Das, S. B. Pooja Rao, Dinesh Babu Jayagopi, "Online peer-to-peer conversations: Automated communication skills evaluation network," Affective Programming and Intellectual Communication Seminars and Demonstrations (ACIWI) 2017 Seventh year.
3. Barron, John M., and the bishop John. "Extensive search, intense recruitment and recruiting costs: new data on recruiting operation for employers." *Economic survey* 23, No. 3 (1985): 363-382.
4. Lei Chen, Ru Zhao, Chee Wee Leong, Blair Lehman, Gary Feng, Mohammed Ehsan Hoque, "Automated video interview decision on a large-scale corpus collected online," 2017 Seventh International Conference on Effective Computation and Intellectual Communication (ACII), pp. 504-509, 2017.
5. Berkay Aydin, Ahmet Alp Kindiroglu, Oya Aran, Lale Akarun, "Automatic personality prediction using random forest regression using audiovisual evidence," Pattern Detection (ICPR) 2016 23rd International Conference on, pp. 37-42, 2016.
6. L. S. Nguyen, M. S. Mast and D. Frauendorfer. Gatica-Perez, "Employ me: Statistical Assumption of Hirability in Nonverbal Behaviour-Based Job Interviews," in *IEEE Digital Transactions*, vol. 16, number 4, pp. 1018-1031, June 2014.
7. W. Wang et al., "Strategic Virtual Team Crowdsourcing: Creating a Team of True Staff in Social media for Crowdsourcing," in *IEEE Mobile Computing Transactions*, vol. 18, number 6, pp. 1419-1432, 1 June 2019.
8. Pascale Petit, Duguet, Pascal Petit, Emmanuel, and ... *Economic and Statistical Annals* (2005): 79-102. "The recruitment of discrimination in the French financial sector: an econometric analysis of findings from field experiments"
9. Vera Brenčić. "Hiring practices for workers, work security and costly search: a vacancy-level study." *Labor Economics* 16, No. 5 (2009): 461-479.
10. Freidlin, Joseph L. Gastwirth and Boris. *Review of Business & Economic Statistics* 18, no. 3 (2000): 315-322. "Changepoint assessments designed to examine recruitment data resulting in employment discrimination cases".

11. Barron, John M., John Bishop, and William C. Dunkelberg. "Employer search: The interviewing and hiring of new employees." *The Review of Economics and Statistics* (1985): 43-52.
12. Barnett, George A., James A. Danowski, Thomas Hugh Feeley and Jordan Stalker. *Examination of Notice* 60, no. 2 (2010): 388-411. "Measuring quality of doctoral contact through network study of patterns in faculty recruitment"
13. Conlin, Michael, Emerson and Patrick M. "Race in recruiting versus retention and advancement: An empirical study of in-company care of NFL players." *Journal of Law, Economics and Organization* 22, No. 1 (2006): 115-136.
14. Papay, John P., and Kraft, Matthew A. "Inefficient recruiting practices harm productive capacity: facts from late teacher recruiting." *Policy Analysis and Management Review* 35, No. 4 (2016): 791-817.
15. Jackson, F J. "Hiring African American male practices in academic leadership roles at American universities and colleges: a job pattern and a differential impact study." *Teachers College Record* 108, No. 2 (2006): 316.

AUTHORS PROFILE



Mr. S Ashok kumar, currently working as Assistant Professor in the Department of Information Technology, Sri Shakthi Institute of Engineering and Technology, Coimbatore. And, Completed Post Graduation in Computer Science and Engineering from Anna University of Technology, Coimbatore, Under Graduation in Information Technology from Sona College of Technology, Salem and have 9 years of teaching experience. Life member of ISTE, and IE.



Saranyadevi S Student, Department of Information Technology, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India.



Santhosh V Student, Department of Information Technology, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India.



Vikranth Krishna E Student, Department of Information Technology, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India.